

修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院 情報理工学研究科 情報・通信工学専攻 博士前期課程		
氏 名	萩原 崇文	学籍番号	1231071
論 文 題 目	スマートデバイスにおいて大量の楽曲から 高速に選曲する手法の提案と評価		
要 旨			
<p>本研究では、スマートデバイスにおいて、画面のタップのリズムや位置によって楽曲を選曲するシステムを提案し、実装した。また、実験により、選曲の正確さや有用性について評価を行った。</p> <p>近年、Android や iOS のような汎用 OS を搭載した音楽プレイヤーやスマートフォンで音楽を聞く機会が増加している。これらの端末は数千曲もの楽曲を保存し再生することができる。しかし、一方で、端末内の楽曲数の増加により、曲やアルバム等の選曲画面からの選曲に時間がかかってしまうという問題が発生する。</p> <p>そこで、本研究では、携帯端末用の直感的な選曲システムを提案した。端末の画面をタップする時のリズムやタップ位置による音程変化の情報を楽曲のボーカルの発音のリズムや音高の変化と比較することにより聴きたい曲を絞り込むシステムを設計および評価した。</p> <p>本研究では、楽曲の音響特徴から楽曲データベース(DB)を作成するシステム、端末の画面から入力を行うシステム、楽曲 DB の楽曲データと入力されたタップデータを比較し選曲するシステムを作成した。</p> <p>予備実験により、入力方法は端末の画面をピアノの鍵盤のように入力する方法とした。また、アルゴリズム評価の有用性について評価を行った。入力実験により、選曲の正確さや有用性について評価を行った。</p> <p>RWC 研究用音楽 DB の 100 曲を対象として比較する際、MIDI データのデータベースを用いた場合は、入力データが正確な場合、100%選曲に成功するが、被験者による入力データでは、最良で 5 位以内に入る確率が 91.7%であった。波形データの DB を用いて選曲した際は、被験者による入力データでは最良で 5 位以内に入る確率は 38.4%であった。また、被験者が入力したデータの音高推移の正しさは 77.8%程度であった。</p> <p>これらの結果より、選曲時に入力と DB を比較するアルゴリズムの改良、また、ユーザーの訓練によって選曲の精度を改善することが可能であると考えられる。</p>			

2013 年度修士論文

スマートデバイスにおいて大量の楽曲から
高速に選曲する手法の提案と評価

電気通信大学大学院 情報理工学研究科
情報・通信工学専攻 コンピュータサイエンスコース
角田 HI* 研究室

指導教員： 角田 博保 (*Kakuda Hiroyasu*)
 中山 泰一 (*Nakayama Yasuichi*)

学籍番号：1231071 / 萩原 崇文 (*Hagihara Takafumi*)

提出日：2014 年 1 月 27 日 (月)

*Human Interface

概要

本研究では，スマートデバイスにおいて，画面のタップのリズムや位置によって楽曲を選曲するシステムを提案し，実装した．また，実験により，選曲の正確さや有用性について評価を行った．

近年，Android や iOS のような汎用 OS を搭載した音楽プレイヤーやスマートフォンで音楽を聞く機会が増加している．これらの端末は数千曲もの楽曲を保存し再生することができる．しかし，一方で，端末内の楽曲数の増加により，曲やアルバム等の選曲画面からの選曲に時間がかかってしまうという問題が発生する．

そこで，本研究では，携帯端末用の直感的な選曲システムを提案した．端末の画面をタップする時のリズムやタップ位置による音程変化の情報を楽曲のボーカルの発音のリズムや音高の変化と比較することにより聴きたい曲を絞り込むシステムを設計および評価した．

本研究では，楽曲の音響特徴から楽曲データベース (DB) を作成するシステム，端末の画面から入力を行うシステム，楽曲 DB の楽曲データと入力されたタップデータを比較し選曲するシステムを作成した．

予備実験により，入力方法は端末の画面をピアノの鍵盤のように入力する方法とした．また，アルゴリズム評価の有用性について評価を行った．入力実験により，選曲の正確さや有用性について評価を行った．

RWC 研究用音楽 DB の 100 曲を対象として比較する際，MIDI データのデータベースを用いた場合は，入力データが正確な場合，100%選曲に成功するが，被験者による入力データでは，最良で 5 位以内に入る確率が 91.7%であった．波形データの DB を用いて選曲した際は，被験者による入力データでは最良で 5 位以内に入る確率は 38.4%であった．また，被験者が入力したデータの音高推移の正しさは 77.8%程度であった．

これらの結果より，選曲時に入力と DB を比較するアルゴリズムの改良，また，ユーザーの訓練によって選曲の精度を改善することが可能であると考えられる．

目次

第1章	はじめに	6
1.1	大容量の携帯音楽プレイヤーの普及と問題点	6
1.2	この問題に関するアプローチ	6
1.3	本研究の目的	6
1.4	本論文の構成	7
第2章	関連研究	8
2.1	リズムカルなタップによる楽曲検索	8
2.2	鼻歌による楽曲検索	8
2.3	アクセント情報を利用したコマンド入力	8
2.4	音響特徴を利用したプレイリストの自動作成	9
2.5	本研究の位置づけ	9
第3章	提案手法	11
3.1	提案手法の概要	11
3.2	入力画面	12
3.3	リズム・音高推移の分析	13
3.4	入力データと楽曲データの比較	14
第4章	実装	18
4.1	実装方法	18
4.2	実装内容	18
第5章	予備調査	21
5.1	調査方法	21
5.2	調査結果と考察	21
第6章	予備実験	23
6.1	実験の目的	23
6.2	画面のタップ操作	23
6.3	実験手順	23
6.4	被験者とインタラクションの割り当て	25
6.5	実験結果	25
6.6	考察	25

第 7 章	入力・楽曲比較システムの評価	28
7.1	評価の目的	28
7.2	比較するデータおよびその対象	28
7.3	評価の結果	29
7.3.1	MIDI から作成した DB との比較結果	29
7.3.2	楽曲波形を解析した DB との比較結果	31
7.4	考察	31
第 8 章	評価実験	33
8.1	実験方法	33
8.2	被験者と楽曲の割り当て	33
8.3	実験結果	34
8.4	考察	35
第 9 章	別アプローチによる再評価	37
9.1	評価方法	37
9.1.1	比較方法の詳細	37
9.2	評価の結果	38
9.3	考察	38
第 10 章	おわりに	44

目 次

2.1	関連研究 [1] より, 楽曲データと入力データの処理の流れ	10
2.2	関連研究 [1] より, ボーカルの強調方法の概要	10
2.3	関連研究 [2] より, タップとアクセントの入力方法	10
3.1	システムの概要図	11
3.2	入力画面 (4 ボタン)	12
3.3	楽曲のメロディから入力を行う方法	12
3.4	楽曲データの処理の流れ	13
3.5	ピッチ検出	14
3.6	楽曲の発音とタップのタイミングを合わせる	15
3.7	タップと発音のタプル生成方法模式図	17
6.1	操作画面:タイプ1	24
6.2	操作画面:タイプ2	24
6.3	操作画面:タイプ3	24
6.4	端末の把持	26
6.5	被験者別および入力インタラクション別のタップ正解率	26
7.1	入力データに対するスコアの順位についての度数分布 (楽曲波形) . .	32
8.1	タップした曲に対するスコアの順位についての度数分布	35

表 目 次

5.1	予備実験に使用した端末のスペック	21
6.1	セッション1の被験者・曲別正解率(単位:%)	27
6.2	セッション2の被験者・曲別正解率(単位:%)	27
7.1	各ボタン数における, ターゲットとなる曲の他曲との比較結果 (「MIDI から生成した楽曲 DB」の場合)	29
7.2	各曲の入力における, 目的の曲が指定順位以内に入った曲数 (%) . .	31
8.1	各曲の入力における, 目的の曲が指定順位以内に入った割合 (%) . .	34
8.2	被験者別平均順位	35
9.1	各曲の入力における, 目的の曲が指定順位以内に入った割合 (MIDI)(%)	40
9.2	各曲の入力における, 目的の曲が指定順位以内に入った割合 (楽曲 波形)(%)	41
9.3	被験者別平均順位 (MIDI)	42
9.4	被験者別平均順位 (楽曲波形)	42
9.5	MIDI から入力した場合において, 目的の曲が指定順位以内に入っ た割合 (%)	42
9.6	LCS による音高推移一致率 (%)	43

第1章 はじめに

1.1 大容量の携帯音楽プレイヤーの普及と問題点

現在, Walkman や iPod Touch のように, Android や iOS のような汎用 OS を搭載した音楽プレイヤーが普及してきている. また, それらの汎用 OS を搭載したスマートフォンも普及してきており, それらの端末で音楽を聞く機会も増加してきている. これらの端末は, 数十 GB もの容量を持っている場合が多く, 非常に多くの楽曲を保存し, 再生することが可能である.

しかしながら, それらの端末は, アルバムやアーティストの一覧から曲を選んで行くものが多い. そのような方式だと, 端末内の楽曲やアーティスト, アルバムが増えて来た場合に, 何度もそのリストをスクロールさせなければならない. また, 画面を凝視して曲名を探す手間がかかったり, 聴きたい曲の名前を忘れてしまうといった問題が発生する.

1.2 この問題に関するアプローチ

汎用 OS を使用している端末では, 文字入力を行うことにより目的の曲を素早く見つけることが可能である. しかし, アルファベットや数字・かな漢字のように複数の文字種が混ざった時に文字入力に手間がかかってしまう. また, 非常に多くの曲が端末に入っている場合はその曲の名前を忘れてしまい検索に時間がかかってしまう問題が発生する.

また, 鼻歌等で楽曲を検索できるサービスも存在するが, 人混みや電車の車内のように, 周囲に人がいたり騒がしかったりすると使えないといった問題も発生する.

1.3 本研究の目的

以上を踏まえて, スマートデバイス上においてのタップ操作によって楽曲を検索する方法を提案する.

本研究では, 端末の画面をタップする時のリズム・強さ・タップ位置による音程変化の指定の情報を楽曲のボーカルの発音のリズム・音の特徴・音高の変化と比較することにより, 利用して聴きたい曲を絞り込み, より素早く選曲するシステムを設計・評価する.

端末で, 音高変化と感情の強さを入力し, 予め作成された楽曲側のデータと比較を行うことにより, 入力データにより近い楽曲を選曲できるようにする.

1.4 本論文の構成

本論文の構成を以下に述べる．第 2 章では，本研究に関連する研究を挙げる．第 3 章では，本研究における提案手法とシステムの詳細について，選曲したい楽曲の入力方法や楽曲 DB の生成方法，入力データと楽曲 DB の比較について説明する．第 4 章では，具体的なシステムに実装について述べる．

第 5 章では予備調査について，第 6 章では予備実験について，主に，選曲したい楽曲の入力方法を決定するための調査や研究について述べる．

第 7 章では，評価実験に先立ち，提案したシステムやアルゴリズムが正しく機能しているかを検証する．第 8 章では，実際の被験者を用いた評価実験の詳細と結論・考察などを述べる．第 9 章では，評価実験の結果に基づいて別のアプローチから再評価を行ったので，その内容と結果について説明する．第 10 章では，本研究の結論と展望について述べる．

第2章 関連研究

この章では、本研究に関連する研究として、タップを利用した楽曲検索、鼻歌を利用した楽曲検索、タップによるコマンド入力時にアクセントを利用する入力方法、音響特徴を利用した自動的なプレイリスト作成について説明する。

2.1 リズミカルなタップによる楽曲検索

池谷ら [3] の研究では、1 個のボタンをリズミカルにタップする事により、音楽を検索するアルゴリズムを提案している。

MIDI から検索用のデータを作成し、ユーザーは探したい曲を打鍵のリズムによって入力する。入力されたデータは、打鍵のタイミング・リズムのみを処理し、打鍵の早さが異なっても問題がないとしている。また、楽曲側のデータとは、前方一致で比較を行う。

実験では、2500 曲中、意図した曲が5 位以内に表示される確率が75%という結果になっている。

2.2 鼻歌による楽曲検索

Bandera ら [1] の研究では、鼻歌により楽曲検索を行うことができる。検索用のデータは楽曲波形から生成している。

図 2.1 の流れのように、楽曲の波形からボーカルの部分を取り出し、楽曲の特徴を示したフィンガープリントを生成する。入力されたハミングに対しても同様にフィンガープリントを作成し、入力されたハミングと楽曲データを比較していく。

楽曲から検索用のデータを作成する際にボーカルを取り出す時、楽器の音の定位が左右にふられていることを利用する。図 2.2 のように、左右のチャンネルの波形のスペクトルを比較し、類似している部分を取り出すことによりボーカルが強調される。

実験の結果、140 曲からポップスやロックの曲を探す場合において、意図した楽曲が5 位以内に表示される確率が45.10%であると示されている。

2.3 アクセント情報を利用したコマンド入力

タップにリズム以外の情報を付加したものとして、石山ら [2] の研究が挙げられる。

音声コマンドの発音のリズムをタップ操作に置き換えて入力する際に、別のコマンド同士で類似したタップ操作が存在する。そのような場合に、音声コマンドを発音するときのアクセントの強弱を同時に入力することにより、複数の類似したコマンドを区別する。

タップ操作とアクセントの入力は、図 2.3 のように携帯端末の側面の 2 つのボタンを利用して行っている。

2.4 音響特徴を利用したプレイリストの自動作成

Rui ら [4] の研究では、様々な楽曲の音色等の音響特徴からシグネチャを作成し、音響特徴が近い楽曲同士で自動的にプレイリストを作成するシステムについて示されている。

2.5 本研究の位置づけ

2.1 節で述べた研究は、楽曲の MIDI に対して検索を行い、タップのリズムのみを選曲に利用していた。しかし、本研究では実際の楽曲の波形を用い、また、それに伴う精度向上のためにタップの強さも利用する。

また、2.2 節の研究では鼻歌を利用していたが、鼻歌は電車の中など、周囲に迷惑をかける可能性や周囲の騒音により正常に入力できない可能性もあるため、それに代わる周囲の状況に左右されない方法として、本研究ではタップ入力を提案する。

2.3 節の研究においては、楽曲ではなく端末操作のためにタップ入力を行っているが、端末側面のボタンを利用しているため、リズムとアクセント以外の情報を入力することが困難である。本研究では、画面のタップ操作を行うことにより、より多くの種類の情報を入力できるようにする。

2.4 節において、単に音色などの音響特徴を用いた照合方法では、特定の曲を探す目的でユーザーがタップ操作によって検索クエリを入力するには不向きであると考えられる。そのため、よりユーザーにとって分かりやすいタップの強さの指標を提案する。

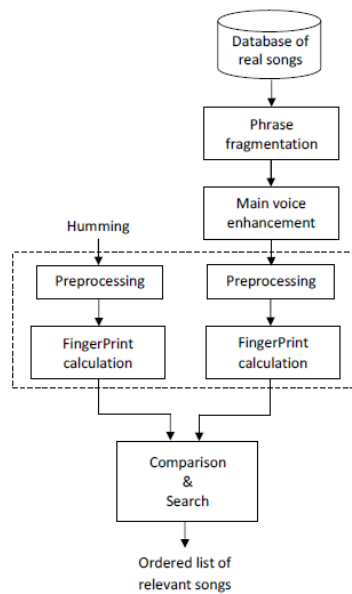


図 2.1: 関連研究 [1] より，楽曲データと入力データの処理の流れ

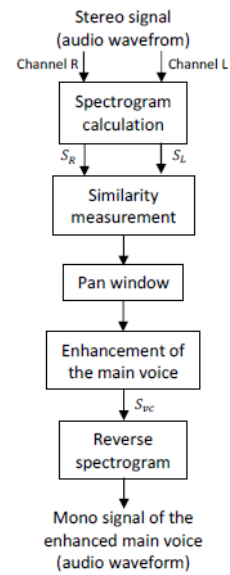


図 2.2: 関連研究 [1] より，ボーカルの強調方法の概要

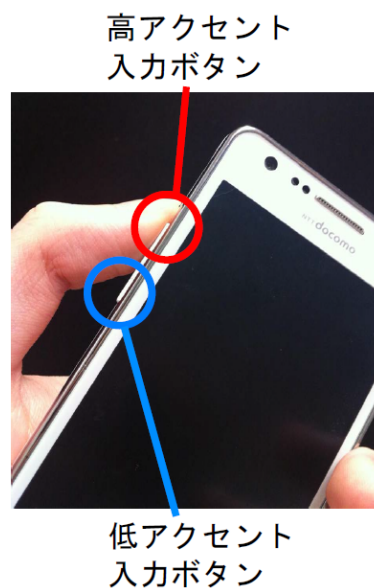


図 2.3: 関連研究 [2] より，タップとアクセントの入力方法

第3章 提案手法

本章では、提案手法とシステムの設計について説明する。

3.1 提案手法の概要

本研究では、スマートフォンを使用してタップ入力により選曲を行うシステムを提案する。端末の画面にボタンが表示される。ユーザーは、聴きたい曲のリズムに合わせて端末の画面に表示されたボタンをタップする。その時、リズム・曲の感情や印象・メロディの「音高推移」を利用する。メロディの音高推移とリズムはボタンをタップする順番やタイミングによって表現する。音高推移とは、楽曲のある音が、前の音よりも低くなったか、変わらないか、あるいは高くなったかといった変化のことを言う。また、曲の感情や印象は、タップの強弱によって表現する。例えば、激しい曲やポジティブな曲では強くタップし、静かな曲やネガティブな曲は弱くタップする。これを「感情レベル」と定義する。

システムの概要図を図 3.1 に示す。

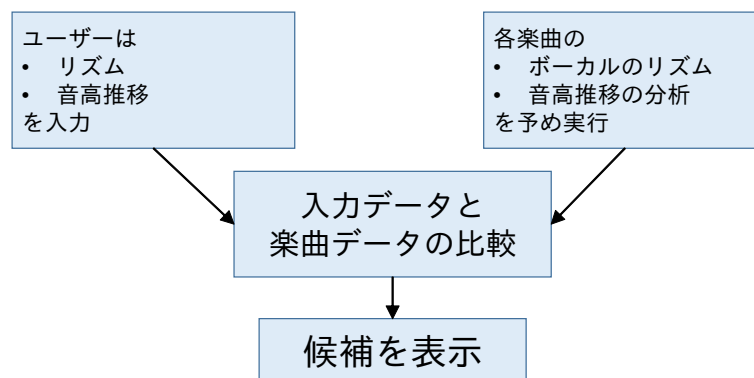


図 3.1: システムの概要図

選曲対象の楽曲に対して、波形データから予めボーカルの発音タイミング・音高推移の分析および楽曲の感情レベルの決定を行ない、選曲用データベース (DB) を自動的に生成しておく。ユーザーはタップ操作で聴きたい曲の情報を入力する。入力されたデータは選曲用 DB と照合し、類似度を比較することにより楽曲を絞り込み、候補を表示する。

3.2 入力画面

ユーザーは、入力画面でリズムに合わせて音高推移に合わせたタップ入力を行う。

リズムに合わせたタップ入力において、ユーザーはビート、またはメロディ(主旋律)に合わせてタップ操作を行うものと考えられる。そこで、ユーザーがどのように画面をタップするのかを調査するため、予備調査(5章)を行ったところ、ユーザーはメロディに合わせてタップ操作を行う傾向があることがわかった。よって、本研究では、メロディに合わせて入力を行うようにした。

6章の予備実験により、楽器の鍵盤を弾くような感覚で入力すると入力が行いやすいと考えられる。また、後述の、タップ入力と楽曲DBの比較システム(3.4節)について、7章で検証したところ、4ボタンの場合が選曲精度が上がることが分かった。また、野口ら[5]の研究により、指先が器用な人は4本指でリズムよくタップを行うことができることが分かった。本研究の入力方法では、指先が器用でない人も4本より少ない指で端末の入力画面をタップすることが出来ると考えられることも鑑みて、使用するボタン数は4つとした。

画面上に、図3.2のように左右に4個のボタンを配置する。

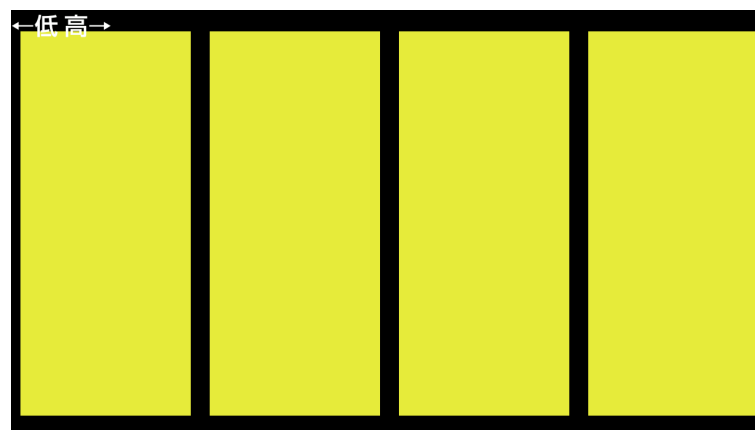


図 3.2: 入力画面 (4 ボタン)

音高推移を入力していく時、図3.3のように、音高が上がる時は、1つ前に押したボタンの右隣をタップする、但し、1つ前が右端のボタンだった場合は同じ右端のボタンを押す。反対に、音高が下がる時は、1つ前に押したボタンの左隣をタップする。但し、1つ前が左端のボタンだった場合は、同じ左端のボタンを押す。1つ前の音と同じ音の場合は、同じボタンをタップする。

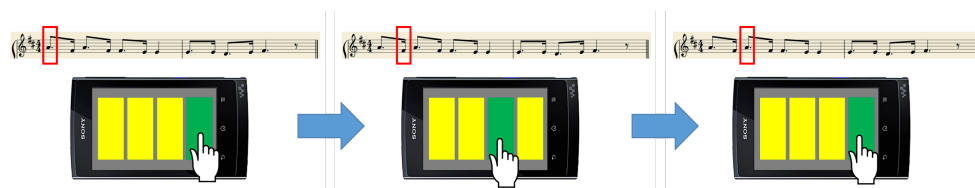


図 3.3: 楽曲のメロディから入力を行う方法

また、入力時に、感情レベルの強さに応じてボタンをタップするときの強弱をつける。これらの操作によって、ボーカルの発音リズムと音高推移を同時に入力する。

3.3 リズム・音高推移の分析

リズム・音高推移を分析する際、図 3.4 に示す流れで楽曲データを処理する。楽曲の波形からボーカルを抽出したあとに、その音声波形のパワーが急峻な部分を検出して発音タイミングを分析し、また、音声波形のピッチの情報を取り出す。

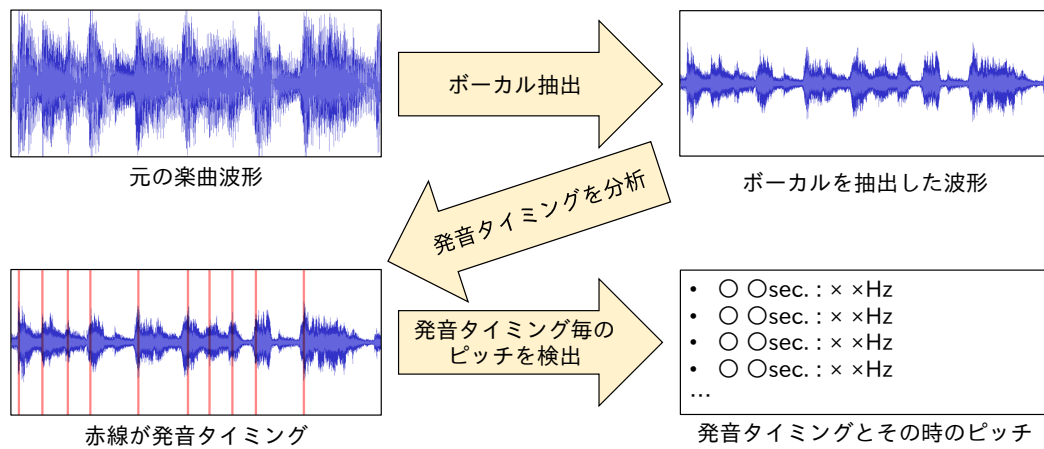


図 3.4: 楽曲データの処理の流れ

ボーカルの抽出には、REPET-SIM[6, 7]を用いる。REPET-SIMは、音響信号から楽器のベース音などの繰り返される音響成分を残し、ボーカルを除去した波形を得ることが出来る。その後、元の波形からボーカルを除去した波形を減算すると、ボーカルが抽出された波形を得ることが出来る。

ここで、RAFIIらの研究[6, 7]よりREPET-SIMの概要について述べる。ある楽曲の波形を、ウィンドウサイズ N の半分のサンプル数 $n = N/2$ ごとに短時間フーリエ変換し、そのスペクトログラムを並べたものを $V(i, j)$ とする。この時、 j は先頭から数えたスペクトログラムのフレーム番号で、 i は各スペクトログラムにおける周波数領域を表す。ここで、類似マトリクスとして、

$$S(j_a, j_b) = \frac{\sum_{i=1}^n V(i, j_a) V(i, j_b)}{\sqrt{\sum_{i=1}^n V(i, j_a)^2} \sqrt{\sum_{i=1}^n V(i, j_b)^2}} \quad (3.1)$$

を定義する。この類似マトリクスにより、 $V(i, j)$ における2点のスペクトルの時刻 j_a, j_b 間におけるスペクトログラムの類似度を求めることが出来る。各フレーム番号 j に対して、類似しているスペクトログラム特徴を持つフレームが k 個あったとすると、その中の l 番目のフレームを $V(i, J_j(l))$ と表すとする。繰り返される音響成分だけが残されたスペクトルを並べた列を $W(i, j)$ とすると、

$$W(i, j) = \text{median}_{l \in [1, k]} \{V(i, J_j(l))\} \quad (3.2)$$

となるので、その列を逆短時間フーリエ変換することによって、ボーカルを取り除いた波形を得ることが出来る。その波形を元の波形から減算すると、ボーカルが抽出された波形を得ることが出来る。

ボーカルを抽出した波形に対して、発音タイミングの分析を行う。波形のうち、先頭からの時刻 P における一定時間 F の範囲の振幅を 2 乗平均した値を W_p とする。ここで、

$$P = (F/2) \times C (0 \leq C \leq 2L/F - 2) \quad (3.3)$$

N 回目の発音タイミング P_N について、

$$W_{P_N} = \max_{a \in [x_{N-1}, x_N]} W_a \quad (3.4)$$

$$x_N = \min(\{x | P_N < x \wedge W_x < W_{P_N} \times \text{threshold}(x, P_N)\}) \quad (3.5)$$

ただし、 $\text{threshold}(x, a)$ は発音タイミング a に対する x でのしきい値となる。

$$\text{threshold}(x, P_N) = \min\{(x - P_N), 0.5\} \quad (3.6)$$

このようにすることで、図 3.5 のように発音のタイミングを検出することが出来ると考えられる。

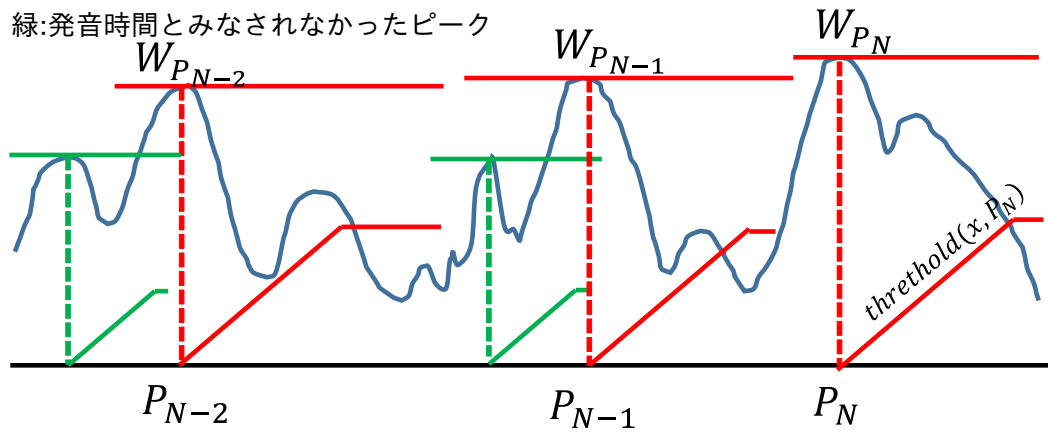


図 3.5: ピッチ検出

ピッチの検出は、発音時刻毎に行う。本研究では、Yin のアルゴリズム [8] によって声のピッチを検出した。この処理によって発音時刻とその時の周波数を関連付けることができる。その情報を元に、入力データと比較するための音高推移のデータやリズムを生成する。

3.4 入力データと楽曲データの比較

入力されたデータと選曲用 DB のデータを基に、類似度を比較する。

曲 m において、 k 回目の発音時刻を $s_{m,k}$ とした時、 k 回発音した時の各発音時刻の列を、

$$\mathbf{S}_m = (s_{m,1}, \dots, s_{m,k}, \dots, s_{m,K}) \quad (3.7)$$

と表す. また, k 回目の発音における周波数を $p_{m,k}$ とした時, 周波数の列を

$$\mathbf{P}_m = (p_{m,1}, \dots, p_{m,k}, \dots, p_{m,K}) \quad (3.8)$$

と表す. l 回目のタップの時刻を s'_l とした時, l 回発音した時の各タップ時刻の列を,

$$\mathbf{S}' = (s'_1, \dots, s'_l, \dots, s'_L) \quad (3.9)$$

と表す. l 回目のタップにおけるボタンを p'_l とした時, ボタンの列を

$$\mathbf{P}' = (p'_1, \dots, p'_l, \dots, p'_L) \quad (3.10)$$

と表す.

入力されたデータは, 選曲用 DB のデータと比較を行う.

各楽曲のスコアを比較し, よりスコアが高い楽曲が, より入力された楽曲のタップ操作に近い楽曲となる.

曲 m のスコアは以下のように求める. まず, 音高推移を比較する曲 m の発音と入力のタップの組合せを作っていく. この時, 楽曲 m のリズムの拍の間隔と入力のタップの拍の間隔が同じであれば, 図 3.6 のように互いの早さが違っても関連付けられるようにする.

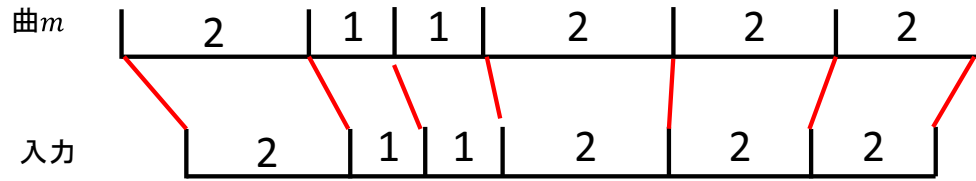


図 3.6: 楽曲の発音とタップのタイミングを合わせる

ここで, その方法について具体的に説明していく.

最初に, タップのリズムパターンとボーカルの発音リズムパターンがより一致すると考えられるパターンの候補を決定する.

曲 m の k 回目の発音と入力の l 回目のタップのタプルを (m, k, l) とする. 例えば, $(m, 1, 1)$ は, 曲 m における 1 回目のタップと 1 回目の発音の組み合わせとなる

ここで, 音高推移を比較するタプル (m, k, l) の直前のタプルを

$$before(m, k, l) = (m, w(k), l - 1) \quad (3.11)$$

とおく.

ここで, $w(k)$ は, $0 \leq i \leq k - 1$ について,

$$\{(s_{m,k} - s_{m,w(k)}) - (s'_l - s'_{l-1})\}^2 \leq \{(s_{m,k} - s_{m,i}) - (s'_l - s'_{l-1})\}^2 \quad (3.12)$$

とする. この式は, 前回のタップ・発音のタプルが今回のタップ・発音のタプルよりもタップの間隔と発音の間隔のズレが小さいことを表す. これによって, タッ

プのリズムパターンと楽曲 m のボーカルの発音リズムパターンがより一致すると考えられるパターンを決定した。

ここで、曲 m の発音とタップのタプル (m, k, l) について、入力の音高推移を $psi(p'_a, p'_b)$ ($a > b$)、曲 m の音高推移を $psm(p_{m,c}, p_{m,d})$ ($c > d$) とする。但し、各 psi および psm は、音高が高くなったら +1、変わらない場合は 0、音高が低くなったら -1 を表す。タプル (m, k, l) におけるスコアは、

$$score(m, k, l) = score\{before(m, k, l)\} + new(p_{m,k}, p_{m,w(k)}, p'_l, p'_{l-1}) \quad (3.13)$$

$$new(e_1, e_2, f_1, f_2) = \begin{cases} 1 & \text{if } psm(e_1, e_2) = psi(f_1, f_2) \\ -1 & \text{if } psm(e_1, e_2) = psi(f_1, f_2) \pm 2 \\ 0 & \text{others} \end{cases} \quad (3.14)$$

と表す。

そして、すべてのタプル (m, k, l) についてスコアを求めたあと、

$$\max_{k \in [1, k]} \{score(x, k, L)\} \quad (3.15)$$

がタップ入力に対する曲 m のスコアとなる。

この方法を図表で表すと、図 3.7 のようになる。各セルは、各 k と l の組合せにおけるスコア (上段) と、前回のタップ (赤実線矢印) とのタップ・発音タイミングずれの差 ($\{(s_{m,k} - s_{m,i}) - (s'_l - s'_{l-1})\}^2$) を表している。各セルから出ている矢印が $before(x_{m,k,l})$ の候補であり、 $\{(s_{m,k} - s_{m,i}) - (s'_l - s'_{l-1})\}^2$ の値がより小さいセルが、そのセルから選択された前回のタップのセル (赤実線) である。最終的に、カラム K の中で、一番スコアが高いものをその曲 m のスコアとする。

このアルゴリズムにより、1つのタップ入力に対し、よりスコアの高い曲がより選曲したい曲に近いものになると考えられる。

		$K - 3$	$K - 2$	$K - 1$	K
	
$l - 2$...	12 1.21	13 1.17	3 1.15	6 1.13
$l - 1$...	5 1.23	12 1.19	14 1.18	12 1.15
l	...	7 1.25	6 1.21	12 1.20	14 1.17
$l + 1$...	7 1.27	5 1.23	5 1.22	13 1.19
$l + 2$...	12 1.29	7 1.25	5 1.23	5 1.21
	

図 3.7: タップと発音のタプル生成方法模式図

第4章 実装

4.1 実装方法

本研究では，スマートデバイスによって入力されたタップのデータを計算機に送信し，計算機内で楽曲データの解析や選曲用データベースの作成，および選曲用 DB のデータと入力データの比較を行う方式とした．

Repet-Sim によりボーカルを抽出した音響信号のピッチ解析は，音響信号解析ライブラリ MARSYAS[9] の AubioYin モジュールを使用した．AubioYin モジュールは，Yin のアルゴリズム [8] によって声のピッチを検出することができる．

楽曲 DB の生成・比較システム実装の環境は以下のとおりである．

OS Microsoft Windows 7 64bit

CPU Intel Core i3 M370

メモリ 8GB

言語 C# および C++/CLI

また，入力端末の環境は以下のとおりである．

端末 (OS) Galaxy SII (Android 2.3)

言語 Java

4.2 実装内容

楽曲 DB の生成・タップとの比較システムにおいては，全体として C# を使用した．また，Android 端末の入力システムは Java で記述した．

まず，楽曲 DB の生成・比較システムのプログラムの構造を以下で述べる．

- MusicTapFinder_Preview1.exe

実際に楽曲の管理追加や管理を行う．MusicMetaDataCreatorLocal オブジェクトを呼び出して楽曲の管理を行う，また，MusicExtractor クラスを呼び出して解析を行う．

- MusicFeaturesManager.dll

CompareToMusicAndTapTimeLine 抽象クラス 曲とタップを実際に比較するクラス (3.4 節の部分) を実装する．

FFMPegWrapper クラス FFmpeg[10] を利用して, wav ファイルや mp3 ファイル等を PCMDDataStructure オブジェクトに変換する.

MusicExtractor クラス 元の楽曲の波形から, 楽曲特徴 (タップのリズム・音高推移) を分析する. 後述の VocalPitchDetector クラスを呼び出し, Repet-Sim によるボーカル分離やピッチの検出を行う.

MusicMetaDataCreatorLocal クラス 楽曲 DB (MusicTrackList オブジェクト) への楽曲追加やリストの解析操作 (MusicExtractor クラスの呼び出し) を行う.

MusicSearch クラス 入力されたタップ情報を元に, 実際にすべての曲に対してスコアの計算を行い, その結果の並べ替え・出力を行う.

MusicTrackList クラス 3 章で説明した, 楽曲 DB がこのオブジェクトにあたる. 各楽曲の基本情報 (曲名・アーティスト名等) や音響特徴を格納した, MusicMetaDataBase オブジェクトのリストを保持する.

MusicMetadataLocal クラス / **MusicMetaDataBase** 抽象クラス

MusicMetaDataBase 抽象クラスにおいて, 各楽曲の基本情報および, 楽曲の特徴情報の保持を行う. 特徴情報は, MelodyFeature オブジェクトのリストとして保持している. 本研究では, 1 曲につき 1 つの特徴情報を保持しているが, 複数箇所 (例: サビや間奏の両方を保持) を保持できるようにした. また, MusicMetadataLocal クラスでは, 計算機本体内の楽曲データの保存先について記述している.

MelodyFeature クラス 楽曲の特徴の一部分を表す. 1 曲につき 1 つのオブジェクトであるが, 先述の通り, 1 曲につき 2 つ以上持てるようにした. 楽曲の一部分の発音のタイミング・周波数の列として, SoundFeaturesAtTime オブジェクトのリストを保持する.

SoundFeaturesAtTime クラス 楽曲のある発音の時刻と周波数の情報のペアを保持している.

- PCMDDataStructure.dll

PCMDDataStructure クラス 楽曲の波形 (PCM) の情報を保持するクラス. 1 サンプルを 32 ビット浮動小数点型で保持する.

- VocalPitchDetect.dll

VocalPitchDetector クラス 楽曲に対して実際に Repet-Sim (3.3 節で説明) を行い, 後述の SignalPorocess オブジェクトを用いてピッチの検出及び結果の返却を行う.

- MarsyasWrapper.dll

以下で説明している VocalPitchDete クラスについて, Marsyas プラグインを呼び出している. Marsyas は C++ のプログラムであるため, VocalPitchDetect.dll は C++/CLI で記述し, C# から C++ のプログラムの呼び出しが行えるようにした.

SignalPorocess クラス 3.3 節で説明した内容のうち, AubioYin を実装している Marsyas プラグインを呼び出し, 楽曲波形から発音タイミング・ピッチを検出するクラス.

また, Android のプログラムの構造は以下のとおりである.

- TapExperiment

MainActivity クラス 楽曲 DB の生成・比較システムが稼働している計算機への接続を行う.

ButtonActivity クラス 実際に計算機へのネットワーク接続を行う. ButtonView から受け取った, タップしたボタンの情報をすぐに計算機へ送信する.

PianoButtonView クラス/**ButtonView** 抽象クラス 画面へのボタンの表示, 及び, タップイベントの受け取りを行う. ButtonAcitivity へタップしたボタンの情報を送る.

第5章 予備調査

ユーザーが端末に聴きたい曲の情報を入力する際、メロディを入力させるかビートを入力させるかを決定するため、ユーザーが音楽のリズムを指でどのように刻むのか調査を行った。

5.1 調査方法

ある楽曲の1フレーズ(15～30秒)を3回聴かせ、その後、その曲に合わせてスマートフォンの画面をタップさせた。タップのタイミングは、ビートを刻む方法でも、メロディを刻む方法でも良いとした。

実験用の端末として、Sony XPERIA NX SO-02D または Sony XPERIA acro HD SO-03D を使用した。それぞれのスペックを表5.1に示す。

各曲ごとに、被験者は最初に1回だけどのような曲なのかを聞いた。その後、最初に流れた曲と同じ曲が流れるので、その曲に合わせて、被験者はリズムやビートなど、各自にとって刻みやすいリズムで端末をタップした。

被験者3人であった。3人に対し同じ10曲で調査を行った。

本調査では、RWC 研究用音楽データベース(ポピュラー音楽)[11, 12]の100曲のうち、NO.8,19,37,40,48,50,63,75,86,97の10曲を使用した。

5.2 調査結果と考察

1名は、10曲中9曲についてメロディを刻み、残り1曲(NO.40)でビートを刻んでいることがわかった。残りの2名は10曲全てでメロディを刻んでいることが分

表 5.1: 予備実験に使用した端末のスペック

機種	SO-02D	SO-03D
サイズ (高さ×幅×厚さ)	128mm × 64mm × 10.6mm	126mm × 66mm × 11.9mm
質量	144g	149g
RAM	1GB	
ディスプレイ (サイズ/解像度)	約 4.3 インチ/1280 × 720	
OS	Android 4.0	

かった．よって，本研究において，入力画面では，ユーザーはメロディに合わせて画面をタップするのが適しているとわかった．

第6章 予備実験

本研究では、どのような方法でボタンを操作すると音高推移やリズムの入力が行いやすいかを検証するため、予備実験を行った。本章では、その詳細と考察について述べる。

6.1 実験の目的

3.2節において入力方法を決める際、画面に横にボタンを並べ、そのボタンをタップしていくが、タップによってリズムや音高推移を入力していくインタラクションは複数考えられる。そこで、どのようなインタラクションを採用すれば音高推移やリズムの入力が行い易いかを検証するため、数種類の入力方法を比較、評価する実験を行った。

6.2 画面のタップ操作

この予備実験において、以下の3種類のインタラクションを提案し、比較した。

タイプ1(図6.1)は、ボタン3つで構成されており、音高が下がったら左のボタンを、変わらなかったら中央のボタンを、音高が上がった場合は右のボタンを押す。

タイプ2(図6.2)は、ボタン2つで構成されており、音高が下がったら左のボタンを、音高が上がった場合は右のボタンを押す。音高が変わらなかったらどちらのボタンを押しても良い。

タイプ3(図6.3)は、ボタン2つで構成されており、音高が下がった、または上がった左のボタンを、音高が変わらなかった場合は右のボタンを押す。

6.3 実験手順

この実験では、以下に示す童謡や歌唱などの単音の曲の中からワンフレーズを抜き出して使用した。被験者が流された音の音高推移に合わせてスマートフォン画面に入力できるか実験を行った。

実験用の端末として、5.1節と同じく Sony XPERIA NX SO-02D または Sony XPERIA acro HD SO-03D を使用した。

被験者には、椅子に座り、図6.4のように端末を横画面で片手で把持するように指示した。また、画面を操作する側の指は複数同時に使用してもよいとした。

各曲ごとに、被験者は最初に1回だけどのような曲なのかを聞いた。その後、最初に流れた曲と同じ曲が流れるので、それに合わせて音高が下がったか、変わら

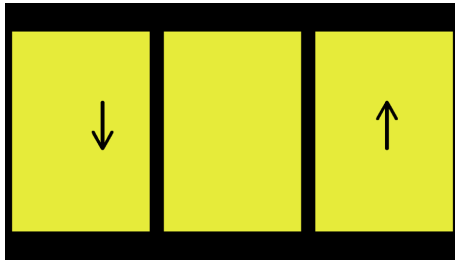


図 6.1: 操作画面:タイプ 1

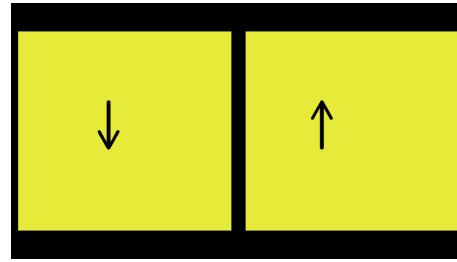


図 6.2: 操作画面:タイプ 2

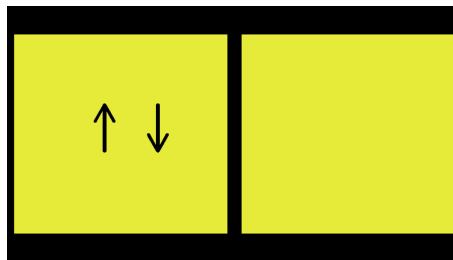


図 6.3: 操作画面:タイプ 3

ないか、あるいは上がったかを 6.2 節で示す端末の画面のボタンを押すことにより入力した。被験者 1 人につき 1 つのインタラクションで実験させ、その割当て方法は 6.4 節で説明している。1 曲につき、入力は 3 回繰り返した。

各被験者は練習を 1 曲行い、その後 10 曲で実験を行った。この手順を 1 セッションとし、1 人 2 セッション行った。

使用した曲は以下のとおり。実験で流れた順に掲載している。

練習 うさぎとかめ

1 どんぐりころころ

2 ふるさと

3 かえるの合唱

4 大きな栗の木の下で

5 夕焼け小焼け

6 アルプス一万尺

7 メリーさんの羊

8 こいのぼり

9 汽車ポッポ

10 かたつむり

6.4 被験者とインタラクションの割り当て

この実験では、本学の角田研究室または赤池研究室に所属している 9 名 (うち女性 1 名) を被験者とした。9 名の被験者の年齢は実験当時 21~25 歳であった。

6.2 節で示される 3 種類のインタラクションに対し、9 名の被験者 A~I から 3 名ずつ割り当てた。被験者は割り当てられたインタラクションで実験を 2 セッション行った。被験者 A~C はタイプ 1 で、D~F はタイプ 2 で、G~I はタイプ 3 で実験を行った。

6.5 実験結果

実験結果について、被験者・曲別の入力正解率を各セッションごとに表 6.1 および表 6.2 に示す。また、入力インタラクション別のタップの正解率を図 6.5 に示す。

図 6.5 より、タイプ 2 が最も正解率が高い。また、タイプ 2 およびタイプ 3 は 1 セッション目より 2 セッション目の方が正解率が高くなっている事がわかる。

6.6 考察

実験結果で示されているように、タイプ 2 の入力インタラクションの時に正解率が一番高くなっている。

タイプ 1 では、音高が下がり続ける、または上がり続ける途中で音高が変わらない部分があった場合、その時だけ別のボタンをタップする操作を要求される。また、音高が下がり続けたあとに音高が上がった時、あるいはその逆の場合に、隣のボタンではなく逆側の端のボタンを押す操作を行わなければならない。そのため、複雑なボタン操作も相まって正解率が下がったと考えられる。

タイプ 3 の場合は、音高が変わらない場合だけ別のボタンを押すという操作方法であるが、特に音高が高くなっていく場合において音高の変化の方向と次にタップするボタンの方向が一致しないため正解率が下がったと考えられる。

一方、タイプ 2 の入力方法の場合、音高が上がっていったら右側をタップし、音高が下がっていったら左側をタップしていくという単純なインタラクションであったため、鍵盤楽器を弾く感覚で直感的に入力することが出来るため、正解率が一番高くなったと考えられる。

よって、タイプ 2 のインタラクションが音高推移やリズムが最も入力しやすいと考えられる。

この実験ではタイプ 2 はボタンが 2 つであったが、そのボタンの数を任意の個数 n にしても入力が行いやすいと考えられる。

この結果を踏まえ、3.2 節では、タイプ 2 の入力に基づく方法を採用した。

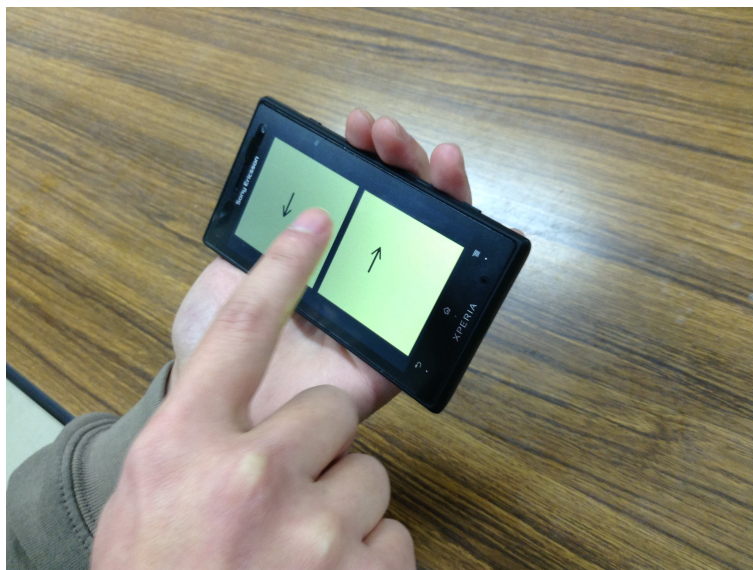


図 6.4: 端末の把持

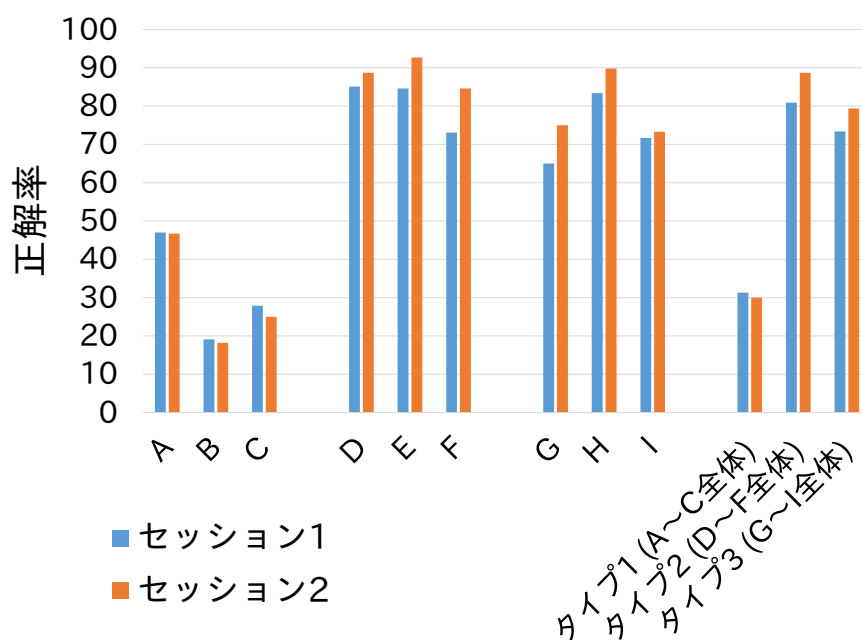


図 6.5: 被験者別および入力インタラクション別のタップ正解率

表 6.1: セッション 1 の被験者・曲別正解率 (単位:%)

タイプ	タイプ 1			タイプ 2			タイプ 3		
被験者 曲	A	B	C	D	E	F	G	H	I
1	57.6	15.2	15.2	93.9	81.8	84.8	66.7	72.7	66.7
2	40.0	26.7	43.3	80.0	76.7	76.7	46.7	83.3	76.7
3	95.2	16.7	54.8	92.9	97.6	90.5	92.9	100.0	100.0
4	42.4	12.1	27.3	81.8	100.0	84.8	75.8	69.7	63.6
5	44.4	8.3	33.3	97.2	86.1	91.7	50.0	88.9	69.4
6	42.2	46.7	22.2	88.9	84.4	35.6	53.3	93.3	86.7
7	38.5	7.7	30.8	89.7	97.4	92.3	74.4	87.2	53.8
8	25.6	25.6	12.8	66.7	66.7	35.9	46.2	64.1	56.4
9	51.1	8.9	28.9	91.1	75.6	64.4	80.0	95.6	82.2
10	33.3	23.1	10.3	69.2	79.5	74.4	64.1	79.5	61.5
全体	47.0	19.1	27.9	85.1	84.6	73.1	65.0	83.4	71.7
タイプ全体	31.3			80.9			73.4		

表 6.2: セッション 2 の被験者・曲別正解率 (単位:%)

タイプ	タイプ 1			タイプ 2			タイプ 3		
被験者 曲	A	B	C	D	E	F	G	H	I
1	51.5	15.2	15.2	97.0	97.0	93.9	90.9	97.0	78.8
2	43.3	23.3	23.3	80.0	90.0	66.7	56.7	80.0	73.3
3	100.0	11.9	38.1	100.0	97.6	97.6	92.9	100.0	95.2
4	33.3	18.2	24.2	90.9	100.0	100.0	90.9	84.8	66.7
5	36.1	16.7	30.6	97.2	100.0	94.4	66.7	94.4	66.7
6	42.2	22.2	24.4	95.6	93.3	77.8	75.6	100.0	88.9
7	53.8	23.1	38.5	89.7	97.4	84.6	69.2	94.9	79.5
8	28.2	15.4	15.4	69.2	69.2	43.6	61.5	64.1	43.6
9	57.8	13.3	22.2	93.3	95.6	100.0	86.7	95.6	88.9
10	20.5	23.1	17.9	74.4	87.2	87.2	59.0	87.2	51.3
全体	46.7	18.2	25.0	88.7	92.7	84.6	75.0	89.8	73.3
タイプ全体	30.0			88.7			79.4		

第7章 入力・楽曲比較システムの 評価

提案した入力・楽曲比較システム(3.4節)について、正確なタップデータを入力した場合に選曲を行うことが出来るか、また、実際の楽曲波形から生成した楽曲DBに対してどれくらい正確に選曲を行うことが出来るか、評価を行った。

7.1 評価の目的

本研究で提案したシステムについて、入力データと楽曲データの比較システムのアルゴリズムが機能しているかどうかを検証するため、入力データと楽曲DBのデータを比較させ、どれくらい正確に選曲を行うことが出来るか評価を行った。

入力するタップデータは、最も正確なデータを作成するためにMIDIから自動生成した仮想的なデータを使用した。

検証の際、リズム・音高推移の分析結果によって影響を受けずに、正確なピッチ情報を取り出すために、MIDIデータから楽曲DBを作成した。また、実際の楽曲波形との比較も行った。

また、端末の入力画面に使用するボタンの個数を検討する材料とするため、複数の入力ボタン数のパターンを試行し、比較を行った。

7.2 比較するデータおよびその対象

本評価では、RWC研究用音楽データベース(ポピュラー音楽)[11, 12]の100曲を使用した。

これらの100曲に対して、音楽データベースに付属するAISTアノテーションのChorus sectionsに基づいて、サビの区間のみを抜き出した。サビを抜き出した楽曲に対し、事前に3.3節で説明した方法に基づいて事前に解析を行い、「楽曲波形を解析した楽曲DB」を作成した。

また、同データベースの各曲についてMIDIデータを用い、先述と同じ区間のデータを抜き出し、MIDIデータの主旋律に当たるパートの発音時刻・ピッチを利用して、「MIDIから生成した楽曲DB」を作成した。

本評価では、人間がタップして生成した入力データの代わりに、正確な入力データを作成した。作成の際、同じMIDIデータの同じ区間・パートのデータを使用した。タップデータはボタンが2~5個の場合の4パターンを生成した。

作成した $100 \times 4 \times = 400$ 通りの入力データは、楽曲 DB の 100 曲と比較してスコアを計算した。

7.3 評価の結果

7.3.1 MIDI から作成した DB との比較結果

「MIDI から生成した楽曲 DB」においてタップデータ 100 通りと比較した結果、すべてのボタン数のすべての楽曲のタップデータにおいて、目的の楽曲が最も高いスコア (1 位) を獲得した。

ここで、各タップ入力によって 1 位をとった曲のスコアが平均からどれだけ大きいかを評価する。すべてのボタン数において、 $M = 100$ 曲中、曲 n をタップした時に、DB 側の曲 m に対するスコアを $R_{n,m}$ とすると、すべてのタップ n において、

$$H_n > 0 \quad (7.1)$$

$$H_n = R_{n,n} - (2\sigma + R_{n,*}^-) \quad (7.2)$$

$$R_{n,*}^- = \frac{\sum_{m=1}^M R_{n,m}}{M} \quad (7.3)$$

$$\sigma = \sqrt{\frac{\sum_{m=1}^{100} (R_{n,m} - R_{n,*}^-)^2}{M - 1}} \quad (7.4)$$

を満たした。式 (7.3) は、あるタップ入力における、DB 中のすべての曲に対するスコアの平均を示す。また、 H_n の値が大きければ大きいほど、平均より値が突出していることを表す。式 (7.1) より、ほとんどのタップのパターンにおいて、目的のスコアが、平均より標準偏差の 2 倍以上を上回っている。これは、目的の曲に対する比較のスコアが、他の曲に対するスコアに対して突出していることを意味する。各曲・各ボタンにおける、 σ , $R_{n,*}^-$, $R_{n,n}$, H_n の値の一覧を表 7.1 に示す。

表 7.1: 各ボタン数における、ターゲットとなる曲の他曲との比較結果
(「MIDI から生成した楽曲 DB」の場合)

ボタン数	2				3				4				5			
n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n
1	5.09	9.27	40.00	16.38	3.40	6.79	40.00	23.01	3.11	6.42	40.00	24.04	2.75	5.98	40.00	25.29
2	7.38	9.14	35.00	9.34	5.91	7.63	35.00	13.84	4.55	6.35	35.00	17.74	3.71	5.70	35.00	19.90
3	4.37	8.46	41.00	19.70	3.41	7.01	41.00	23.58	2.84	6.18	41.00	25.80	2.31	5.51	41.00	27.66
4	1.13	6.23	55.00	41.40	0.94	5.87	55.00	42.31	0.92	5.83	55.00	42.42	0.92	5.83	55.00	42.42
5	0.90	6.34	60.00	46.42	0.82	6.18	60.00	46.81	0.78	6.11	60.00	47.00	0.76	6.08	60.00	47.07
6	11.23	5.13	22.00	0.51	9.44	4.59	22.00	3.38	7.88	4.23	22.00	5.67	7.04	4.02	22.00	6.92
7	8.35	8.26	31.00	6.13	6.43	6.87	31.00	10.84	5.82	6.43	31.00	12.33	5.12	5.94	31.00	13.99
8	8.99	7.24	29.00	5.53	8.80	7.12	29.00	5.95	7.90	6.60	29.00	7.90	7.90	6.60	29.00	7.90
9	10.64	2.56	16.00	0.24	8.69	2.51	16.00	2.28	6.89	2.55	16.00	4.02	6.89	2.55	16.00	4.02
10	3.93	7.21	39.00	20.65	2.70	5.85	39.00	24.59	2.23	5.40	39.00	25.98	1.70	4.89	39.00	27.52
11	6.13	8.66	36.00	12.54	3.34	5.81	36.00	21.03	2.88	5.44	36.00	22.25	2.88	5.44	36.00	22.25
12	7.95	8.38	32.00	7.29	5.44	6.43	32.00	13.71	4.96	6.09	32.00	14.87	4.96	6.09	32.00	14.87
13	10.77	4.19	20.00	0.85	8.89	3.80	20.00	3.51	7.33	3.50	20.00	5.68	7.33	3.50	20.00	5.68
14	9.79	4.51	21.00	2.19	6.95	3.79	21.00	6.47	6.95	3.79	21.00	6.47	6.44	3.67	21.00	7.22
15	1.74	6.71	50.00	34.84	1.22	5.66	50.00	37.45	1.22	5.66	50.00	37.45	1.22	5.66	50.00	37.45
16	7.51	10.59	37.00	8.31	6.05	8.82	37.00	13.30	5.44	8.09	37.00	15.38	4.71	7.29	37.00	17.71
17	1.77	7.61	54.00	37.01	1.35	6.46	54.00	39.72	1.22	6.17	54.00	40.45	1.09	5.91	54.00	41.09
18	7.37	9.09	35.00	9.45	6.13	7.83	35.00	13.22	5.56	7.25	35.00	14.95	4.79	6.54	35.00	17.13

次のページへ続く

前のページから続く (表 7.1)

ボタン数	2				3				4				5			
n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n
19	5.41	10.20	41.00	15.20	4.29	8.31	41.00	20.08	3.24	6.77	41.00	24.23	2.95	6.36	41.00	25.34
20	7.86	8.39	32.00	7.37	6.31	7.15	32.00	11.38	5.20	6.35	32.00	14.09	4.63	5.99	32.00	15.38
21	3.73	8.93	43.00	21.41	2.87	7.22	43.00	25.69	2.54	6.65	43.00	27.16	2.00	5.81	43.00	29.38
22	4.70	7.53	38.00	18.24	3.47	6.26	38.00	22.01	2.66	5.55	38.00	24.23	2.66	5.55	38.00	24.23
23	6.16	9.25	38.00	13.34	4.56	7.49	38.00	18.47	3.77	6.65	38.00	20.93	3.35	6.22	38.00	22.20
24	1.13	6.13	55.00	41.61	1.00	5.87	55.00	42.26	0.95	5.80	55.00	42.45	0.93	5.77	55.00	42.52
25	7.58	7.10	30.00	8.22	5.16	5.65	30.00	13.54	3.98	4.92	30.00	16.19	3.71	4.75	30.00	16.80
26	8.98	3.76	19.00	2.50	7.35	3.71	19.00	4.23	6.28	3.65	19.00	5.42	6.28	3.65	19.00	5.42
27	8.39	8.84	32.00	5.92	7.28	7.87	32.00	8.98	6.64	7.35	32.00	10.66	6.03	6.81	32.00	12.34
28	5.47	8.55	38.00	15.42	4.15	7.07	38.00	19.70	3.60	6.54	38.00	21.32	2.92	5.98	38.00	23.11
29	9.75	6.27	25.00	2.71	6.90	5.36	25.00	7.38	4.71	4.64	25.00	11.01	4.71	4.64	25.00	11.01
30	0.94	9.35	94.00	74.35	0.94	9.35	94.00	74.35	0.94	9.35	94.00	74.35	0.94	9.35	94.00	74.35
31	10.61	1.78	15.00	0.84	9.67	1.79	15.00	1.74	9.67	1.79	15.00	1.74	7.97	1.87	15.00	3.29
32	3.97	7.49	40.00	21.05	3.07	6.27	40.00	24.38	2.78	5.91	40.00	25.40	2.78	5.91	40.00	25.40
33	8.52	8.44	31.00	5.59	7.88	8.00	31.00	7.12	7.40	7.68	31.00	8.25	6.57	7.10	31.00	10.23
34	9.99	6.20	25.00	2.60	6.15	4.69	25.00	9.47	5.22	4.40	25.00	10.97	4.37	4.14	25.00	12.35
35	3.41	6.10	38.00	22.38	2.65	5.37	38.00	24.60	2.31	5.08	38.00	25.52	2.16	4.94	38.00	25.96
36	10.29	4.62	21.00	1.48	9.64	4.44	21.00	2.48	9.10	4.31	21.00	3.28	9.10	4.31	21.00	3.28
37	9.83	6.96	27.00	3.24	6.66	5.51	27.00	9.31	6.27	5.34	27.00	10.05	5.38	4.89	27.00	11.84
38	9.74	7.82	29.00	3.63	7.77	6.73	29.00	7.76	7.40	6.50	29.00	8.60	7.40	6.50	29.00	8.60
39	2.43	7.87	48.00	29.83	1.79	6.45	48.00	33.32	1.46	5.90	48.00	34.73	1.46	5.90	48.00	34.73
40	3.00	7.44	43.00	25.11	2.26	6.18	43.00	28.38	2.07	5.88	43.00	29.18	1.99	5.76	43.00	29.48
41	10.31	8.08	29.00	2.52	6.31	5.81	29.00	11.07	5.69	5.47	29.00	12.36	5.69	5.47	29.00	12.36
42	5.87	8.24	36.00	13.64	4.37	6.65	36.00	18.34	3.01	5.40	36.00	22.18	2.70	5.20	36.00	22.91
43	6.60	3.46	19.00	5.48	5.51	3.40	19.00	6.68	5.51	3.40	19.00	6.68	5.51	3.40	19.00	6.68
44	9.10	8.11	30.00	4.69	7.23	6.80	30.00	9.18	6.76	6.50	30.00	10.24	5.53	5.72	30.00	13.03
45	8.99	6.00	26.00	5.01	8.30	5.68	26.00	6.35	7.32	5.29	26.00	8.10	6.38	4.91	26.00	9.80
46	3.77	10.01	44.00	20.21	3.09	8.49	44.00	23.94	2.95	8.16	44.00	24.72	2.52	7.34	44.00	26.80
47	2.36	7.49	47.00	29.66	1.90	6.50	47.00	32.11	1.46	5.74	47.00	34.06	1.25	5.42	47.00	34.92
48	4.88	8.40	39.00	17.33	3.62	6.80	39.00	21.79	2.36	5.43	39.00	25.79	2.36	5.43	39.00	25.79
49	4.36	8.51	41.00	19.62	3.37	6.98	41.00	23.67	2.84	6.30	41.00	25.56	2.84	6.30	41.00	25.56
50	10.16	5.68	24.00	2.48	7.23	4.84	24.00	7.10	6.01	4.50	24.00	9.00	6.01	4.50	24.00	9.00
51	10.08	8.92	30.00	2.09	7.00	6.79	30.00	9.42	5.50	5.83	30.00	12.84	4.36	5.13	30.00	15.38
52	9.01	8.02	30.00	4.96	7.88	7.20	30.00	7.72	6.97	6.52	30.00	9.99	5.91	5.88	30.00	12.32
53	11.93	6.25	24.00	-0.44	8.09	5.03	24.00	5.85	6.92	4.74	24.00	7.60	6.92	4.74	24.00	7.60
54	10.19	7.11	27.00	2.59	7.41	5.71	27.00	8.18	6.36	5.23	27.00	10.18	5.97	5.03	27.00	10.96
55	11.04	5.41	23.00	1.13	8.38	4.65	23.00	5.32	5.99	4.04	23.00	8.93	5.04	3.83	23.00	10.30
56	6.19	7.95	35.00	12.91	4.65	6.58	35.00	17.20	4.37	6.33	35.00	17.98	4.37	6.33	35.00	17.98
57	8.98	4.30	20.00	2.42	6.79	3.94	20.00	5.33	6.30	3.85	20.00	6.00	5.23	3.62	20.00	7.52
58	10.47	9.83	31.00	0.87	7.58	7.62	31.00	8.17	7.02	7.19	31.00	9.60	5.49	6.21	31.00	13.08
59	9.69	11.02	34.00	2.27	7.02	8.40	34.00	10.19	4.96	6.61	34.00	15.82	4.71	6.45	34.00	16.40
60	11.05	5.52	23.00	0.92	8.84	5.00	23.00	4.17	7.04	4.56	23.00	6.84	6.03	4.29	23.00	8.38
61	5.59	7.28	34.00	13.84	4.79	6.53	34.00	16.16	4.64	6.42	34.00	16.52	3.78	5.73	34.00	18.76
62	7.68	6.59	29.00	8.13	5.76	5.62	29.00	12.00	4.51	5.14	29.00	14.20	3.86	4.78	29.00	15.57
63	9.13	6.01	26.00	4.85	6.29	4.85	26.00	10.02	5.82	4.71	26.00	10.75	5.35	4.58	26.00	11.50
64	5.11	7.13	35.00	15.64	3.24	5.58	35.00	20.61	3.24	5.58	35.00	20.61	3.02	5.44	35.00	21.10
65	4.20	9.12	42.00	19.55	3.11	7.21	42.00	24.48	2.79	6.72	42.00	25.76	2.51	6.26	42.00	26.96
66	7.13	8.48	34.00	9.91	5.55	7.00	34.00	14.45	4.68	6.24	34.00	16.84	4.37	5.96	34.00	17.71
67	10.89	5.84	24.00	1.43	7.47	4.80	24.00	6.93	5.96	4.34	24.00	9.36	5.04	4.14	24.00	10.68
68	8.97	7.34	29.00	5.35	5.74	5.60	29.00	12.05	4.94	5.22	29.00	13.62	4.94	5.22	29.00	13.62
69	8.91	6.55	27.00	4.98	6.85	5.59	27.00	8.98	5.46	4.98	27.00	11.58	4.56	4.64	27.00	13.16
70	5.97	10.04	39.00	12.95	4.44	8.00	39.00	18.56	3.77	7.13	39.00	20.97	3.11	6.39	39.00	23.11
71	11.09	3.84	19.00	0.22	9.07	3.47	19.00	2.99	6.61	3.31	19.00	5.78	6.26	3.31	19.00	6.12
72	8.27	8.17	31.00	6.40	5.45	6.15	31.00	13.24	5.09	5.92	31.00	14.07	3.99	5.26	31.00	16.49
73	6.29	9.14	37.00	12.44	4.93	7.54	37.00	16.98	4.12	6.61	37.00	19.67	3.50	5.95	37.00	21.60
74	10.32	3.10	17.00	0.48	7.21	2.97	17.00	3.86	6.81	2.97	17.00	4.25	6.81	2.97	17.00	4.25
75	10.78	7.51	27.00	1.21	9.29	6.81	27.00	4.10	9.29	6.81	27.00	4.10	8.84	6.59	27.00	4.98
76	10.71	7.47	27.00	1.36	7.80	6.03	27.00	7.14	6.67	5.57	27.00	9.19	5.50	5.08	27.00	11.34
77	10.57	5.76	24.00	1.91	9.05	5.26	24.00	4.43	8.00	4.91	24.00	6.18	6.79	4.53	24.00	8.15
78	4.71	9.01	41.00	18.26	3.70	7.48	41.00	22.33	3.30	6.90	41.00	23.89	2.86	6.34	41.00	25.46
79	10.17	6.59	26.00	2.64	6.75	5.17	26.00	8.91	6.42	5.03	26.00	9.52	5.95	4.85	26.00	10.36
80	7.89	8.80	33.00	7.52	6.34	7.34	33.00	11.99	6.11	7.14	33.00	12.60	5.29	6.44	33.00	14.82
81	7.06	1.38	10.00	0.19	5.74	1.49	10.00	1.27	5.21	1.59	10.00	1.61	5.21	1.59	10.00	1.61
82	4.05	8.75	42.00	20.44	3.51	7.79	42.00	22.90	2.73	6.62	42.00	26.03	2.63	6.46	42.00	26.44
83	9.80	6.87	27.00	3.46	7.57	5.79	27.00	7.85	6.67	5.40	27.00	9.54	5.80	5.00	27.00	11.20
84	6.97	8.67	35.00	10.70	5.57	7.28	35.00	14.88	5.19	6.91	35.00	15.99	5.19	6.91	35.00	15.99

次のページへ続く

前のページから続く (表 7.1)

ボタン数	2				3				4				5			
n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n	σ	$R_{n,*}$	$R_{n,n}$	H_n
85	11.11	7.00	26.00	0.90	8.40	5.79	26.00	6.02	8.10	5.66	26.00	6.57	7.37	5.36	26.00	7.92
86	2.07	7.21	49.00	32.51	1.74	6.45	49.00	34.37	1.46	5.92	49.00	35.71	1.24	5.57	49.00	36.62
87	9.80	2.73	16.00	0.74	8.20	2.68	16.00	2.44	8.20	2.68	16.00	2.44	7.61	2.68	16.00	3.03
88	2.14	8.19	51.00	32.47	1.81	7.26	51.00	34.68	1.68	6.92	51.00	35.48	1.56	6.64	51.00	36.16
89	8.19	8.55	32.00	6.72	5.93	6.68	32.00	12.72	4.61	5.76	32.00	15.87	4.61	5.76	32.00	15.87
90	7.70	8.62	33.00	8.06	4.80	6.12	33.00	15.95	3.59	5.22	33.00	18.98	3.59	5.22	33.00	18.98
91	5.37	8.23	38.00	16.17	4.13	6.77	38.00	20.34	3.63	6.25	38.00	21.87	3.42	6.05	38.00	22.49
92	11.45	6.04	24.00	0.46	8.81	5.28	24.00	4.63	6.61	4.80	24.00	7.79	5.85	4.57	24.00	9.00
93	10.78	2.61	16.00	0.01	8.77	2.62	16.00	1.99	7.46	2.64	16.00	3.27	7.46	2.64	16.00	3.27
94	1.32	7.94	60.00	42.80	1.17	7.26	60.00	44.31	1.06	6.86	60.00	45.22	0.96	6.63	60.00	45.78
95	10.08	2.12	15.00	0.67	8.63	2.13	15.00	2.10	8.63	2.13	15.00	2.10	7.55	2.13	15.00	3.19
96	4.33	9.30	42.00	19.08	3.81	8.36	42.00	21.47	3.55	7.91	42.00	22.63	3.32	7.48	42.00	23.73
97	2.70	7.82	46.00	27.66	2.21	6.74	46.00	30.31	1.81	5.96	46.00	32.28	1.64	5.67	46.00	33.03
98	6.58	5.90	29.00	10.62	4.54	5.01	29.00	14.45	4.11	4.74	29.00	15.42	3.18	4.28	29.00	17.27
99	3.99	9.49	43.00	20.03	3.15	7.80	43.00	24.25	3.03	7.60	43.00	24.78	2.71	7.04	43.00	26.22
100	8.12	4.87	23.00	5.13	6.49	4.46	23.00	7.58	6.09	4.33	23.00	8.26	5.17	4.04	23.00	9.75
平均				11.79				15.68				17.26				18.23

7.3.2 楽曲波形を解析した DB との比較結果

「楽曲波形を解析した楽曲 DB」においてタップデータ 100 通りと比較した結果、各ボタン数において、目的の曲が指定順位以内に入った割合は表 7.2 のとおり。また、入力によって選曲しようとした曲が何位になったかの分布を図 7.1 に示す。

表 7.2: 各曲の入力における、目的の曲が指定順位以内に入った曲数 (%)

ボタン数	1 位	5 位以内	10 位以内	20 位以内	30 位以内
2	2.0	10.0	20.0	40.0	54.0
3	2.0	18.0	24.0	40.0	52.0
4	3.0	17.0	26.0	39.0	47.0
5	2.0	19.0	24.0	42.0	48.0

7.4 考察

評価の結果より、自動生成したタップリズムによって正確なリズム・ピッチの楽曲 DB から楽曲を探すことはできている。また、1 つの入力における、他の曲との比較結果を含めたスコアの平均に対して、目的の曲に対するスコアの差が大きくなっている。

楽曲の音響信号から直接楽曲 DB を生成した場合、では、図 7.1 より、4 ボタンの場合にタップ入力で選曲しようとした曲が 10 位以内に入った割合が一番高かったことがわかった。しかしながら、比較の精度は全体的に劣る結果になっている。アルゴリズムが誤っている場合、または、正しいタイミングおよびその時のピッチを検出しているケース以外に、誤った発音タイミングを検出してしまった場合、お

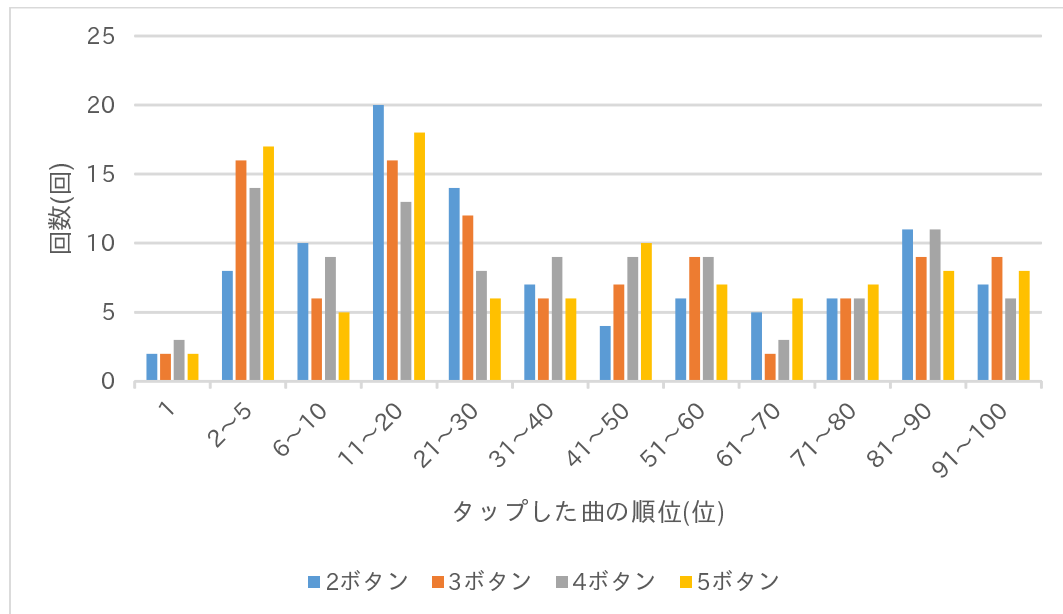


図 7.1: 入力データに対するスコアの順位についての度数分布 (楽曲波形)

よび、正しいタイミングで検出できたとしても、誤ったピッチを検出した場合があると考えられる。

第8章 評価実験

本研究において提案した，入力タップデータと楽曲DBの比較方法，およびユーザーの楽曲のタップ入力の方法に基づき，ユーザーの実際にタップ操作により目的の楽曲をどれだけ正確に選曲することが出来るかを評価する．

8.1 実験方法

本実験では，RWC 研究用音楽データベース (ポピュラー音楽)[11, 12] の 100 曲を使用した．

これらの 100 曲に対して，同音楽データベースに付属する AIST アノテーションの Chorus sections に基づいて，楽曲のサビの先頭からそのサビの終わりの部分までを抜き出した．サビを抜き出した楽曲に対し，事前に 3.3 節で説明した方法に基づいて事前に解析を行い，楽曲 DB を作成した．

タップデータを作成する際は，再生される楽曲に合わせてユーザーが端末のボタンをタップしていく方法をとった．被験者に入力させる際に聞かせる楽曲は，音楽データベースの 100 曲に対し，AIST アノテーションの Chorus sections に基づいて，楽曲のサビの先頭からそのサビの終わりの部分までを抜き出したものである．

タップの入力は，提案した入力画面 (3.2) に基づいて行う．また，端末は予備実験と同様に，片手で端末を把持し，逆側の手の任意の本数の指で画面をタップさせた．得られたタップのデータを，3.4 節において提案する方法に基づいて楽曲 DB と比較を行い，目的の楽曲 (被験者にタップ入力を行わせた時に聴かせた曲) のスコアが楽曲 DB の 100 曲中，何位になるのかを検証した．

被験者は，1 曲につき任意の回数だけ曲を流し，その曲での入力操作に慣れるまで練習を行った，その後，本番として，曲に合わせて入力を行う操作を 3 回行った．

8.2 被験者と楽曲の割り当て

この実験では，本学の角田研究室または赤池研究室に所属している 8 名 (うち女性 1 名) を被験者とした．8 名の被験者の年齢は実験当時 21～24 歳であった．また，被験者の一部は，予備実験 (6 章) の被験者と重複している．

被験者 1 人につき，40 曲のタップデータを入力させた．入力は 1 セッション 10 曲で 4 セッションに分けた．1 人目を 1～40 曲目，2 人目を 41～80 曲目，3 人目を 81～100 曲目及び 1～20 曲目…と割り当てた．

8.3 実験結果

この実験により $8 \times 40 \times 3 = 960$ 個のデータを取得しようと試みた。但し、データの取得に失敗した試行もあったため、実際に得られたデータは 959 個となった。

7 章で作成した「楽曲波形を解析した楽曲 DB」および「MIDI から生成した楽曲 DB」において、被験者のタップ入力から得られた 959 個のデータと比較したところ、目的の曲が指定順位以内に入った割合は表 8.1 のとおり。また、入力によって選曲しようとした曲が何位になったかの分布を図 8.1 に示す。「楽曲波形を解析した楽曲 DB」での全入力に対する選曲したい曲の平均順位は 41.88 位 ($\sigma = 17.89$)、同じく、「MIDI から生成した楽曲 DB」では、27.89 位 ($\sigma = 22.29$) となった。

表 8.1: 各曲の入力における、目的の曲が指定順位以内に入った割合 (%)

被験者	楽曲 DB 作成方法	1 位	5 位以内	10 位以内	20 位以内	30 位以内
1	楽曲波形を解析	0.8	5.8	10.0	22.5	32.5
	MIDI から	5.8	19.2	40.0	69.2	78.3
2	楽曲波形を解析	5.0	9.1	14.2	28.3	41.7
	MIDI から	0.8	4.2	15.0	43.3	51.7
3	楽曲波形を解析	2.5	7.6	11.0	26.9	38.7
	MIDI から	0.8	35.3	49.6	70.6	78.1
4	楽曲波形を解析	0.8	7.5	10.8	25.0	40.0
	MIDI から	4.2	11.7	28.3	53.3	65.8
5	楽曲波形を解析	5.0	8.3	10.8	29.1	44.1
	MIDI から	1.7	15.0	35.0	64.2	69.2
6	楽曲波形を解析	3.3	12.5	20.0	35.0	50.8
	MIDI から	1.7	20.0	43.3	64.1	70.8
7	楽曲波形を解析	5.8	14.2	23.3	35.8	47.5
	MIDI から	0.0	2.5	22.5	53.3	61.7
8	楽曲波形を解析	1.7	4.2	14.2	28.3	37.5
	MIDI から	4.2	21.7	47.5	68.3	72.5
平均	楽曲波形を解析	3.1	8.6	14.2	28.9	41.6
	MIDI から	2.4	16.1	35.1	60.8	68.5

楽曲波形から楽曲 DB を作成した場合、タップ入力から聴きたい曲のスコアを 5 位以内にすることができた割合がわずかに 8.7%、であった。また、MIDI から作成した楽曲 DB と比較した場合は 16.1% となった。目的の曲のスコアが 10 位以内・20 位以内・30 位以内についても、同様に MIDI から作成した楽曲 DB の方が楽曲波形よりも割合が大きくなっていた。また、多くの曲で、MIDI から作成した DB で選曲操作をした時が、楽曲波形から作成した DB で選曲操作をした時よりも順位が良くなることも分かった。

ここで、各被験者別における平均順位を表 8.2 に示す。 $P < 0.05$ より、楽曲波形による DB と MIDI による DB による結果に有意差が見られた。

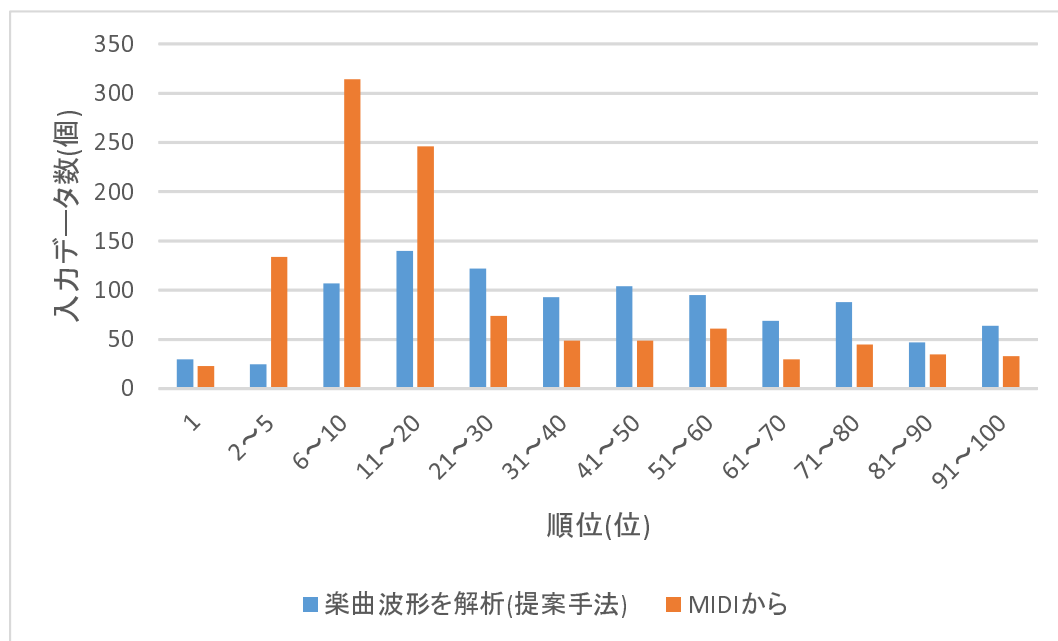


図 8.1: タップした曲に対するスコアの順位についての度数分布

表 8.2: 被験者別平均順位

被験者	楽曲波形	MIDI
1	45.22	23.34
2	41.33	33.60
3	44.80	20.55
4	42.83	30.16
5	41.30	29.60
6	38.93	25.35
7	36.64	31.67
8	46.27	25.12
平均	42.17	27.42
σ^2	10.86	20.28
P	5.61×10^{-4}	

8.4 考察

実験結果より、楽曲波形から DB を作成した場合は、MIDI から DB を作成した場合に比べて順位が低下していた。

これは、MIDI の場合は演奏情報を正確に記録しているため、正確な発音タイミングやピッチを取得することができたためと考えられる。それに対して、7 章でも触れたが、楽曲の波形から Repet-SIM や Yin などにより発音タイミングやピッチを抜き出して作成した楽曲 DB では、誤った発音タイミングやピッチを取得してしまい、選曲の精度が落ちてしまったと考えられる。

また，各被験者別においては，MIDI による DB では被験者によって検索精度にばらつきが生じていることがわかった．これは，被験者によって，音高推移を聞き分ける能力やタップする能力に差があるためであると考えられる．

聴きたい曲の順位が低くなってしまった原因，および，被験者による順位のバラつきについてその原因を調べるため，別アプローチによる再評価 (9 章) においてその原因を調べることにした．

第9章 別アプローチによる再評価

評価実験 (8 章) において，選曲時に聴きたい曲の順位が低くなってしまった原因，および，被験者によって結果にばらつきが生じてしまった原因を調べるため，評価を行った．

9.1 評価方法

評価実験の時に取得した被験者 8 名分のタップ入力データを用いて，以下の条件に変更して選曲したい曲の順位がどのようになるかを検証した．

- タップ時の速度が聴きたい楽曲と等速である前提とする
- タップのリズムのずれのみを比較する，あるいは入力された音高推移の相違のみを比較する，リズム・音高推移の両方を考慮する場合の 3 パターンで比較を行う

9.1.1 比較方法の詳細

3.4 節に示された，DB の曲及び入力データのタップについての各発音時刻・音高推移・周波数の定義について，タップ入力のタップのリズム s'_a と入力のリズムを $s_{m,b}$ の組合せを以下のように定義する．

$$\text{pair}_T(a, h) = b \text{ such that } \min\{((s'_a - s'_h) - (s_{m,b} - s_{m,1}))^2\} \quad (9.1)$$

$$(1 \leq h \leq a \leq L, 1 \leq b \leq K)$$

$$\text{pair}_M(a, i) = b \text{ such that } \min\{((s'_a - s'_1) - (s_{m,b} - s_{m,i}))^2\} \quad (9.2)$$

$$(1 \leq a \leq L, 1 \leq i \leq b \leq K)$$

(1) タップのリズムのズレのみをスコアの計算に用いる場合，曲 m のスコアは，

$$\max\left\{\max_{h' \in [1, L]} T_{m, h'}, \max_{i' \in [1, K]} M_{m, i'}\right\} \quad (9.3)$$

となる．

$$T_{m, h} = \sum_{a=h}^L \text{new2}(s_{m, \text{pair}_T(a, h)}, s_{m, \text{pair}_T(a-1, h)}, s'_a, s'_{a-1}) \quad (9.4)$$

$$M_{m, i} = \sum_{a=1}^L \text{new2}(s_{m, \text{pair}_M(a, i)}, s_{m, \text{pair}_M(a-1, i)}, s'_a, s'_{a-1}) \quad (9.5)$$

$$\text{new2}(e_1, e_2, f_1, f_2) = \frac{1}{\text{Abs}((e_1 - e_2) - (f_1 - f_2)) + 1} \quad (9.6)$$

$$\text{Abs}(c) = \max\{c, -c\} \quad (9.7)$$

- (2) 入力された音高推移の相違のみをスコアの計算に用いる場合、曲 m のスコアは、

$$\max\left\{\max_{h' \in [1, L]} T'_{m, h'}, \max_{i' \in [1, K]} M'_{m, i'}\right\} \quad (9.8)$$

となる。ただし、 $new(e_1, e_2, f_1, f_2)$ は式 (3.14) に定める。

$$T'_{m, h} = \sum_{a=h}^L new(p_{m, pair_T(a, h)}, p_{m, pair_T(a-1, h)}, p'_a, p'_{a-1}) \quad (9.9)$$

$$M'_{m, i} = \sum_{a=1}^L new(p_{m, pair_M(a, i)}, p_{m, pair_M(a-1, i)}, p'_a, p'_{a-1}) \quad (9.10)$$

- (3) リズムのズレと音高推移の両方をスコアの計算に用いる場合、曲 m のスコアは、

$$\max\left\{\max_{h' \in [1, L]} T''_{m, h'}, \max_{i' \in [1, K]} M''_{m, i'}\right\} \quad (9.11)$$

となる。

$$T''_{m, h} = T_{m, h} + T'_{m, h} \quad (9.12)$$

$$M''_{m, i} = M_{m, i} + M'_{m, i} \quad (9.13)$$

9.2 評価の結果

MIDI および楽曲波形によって作成した DB の場合、タップのリズムのズレのみを考慮した場合と、入力された音高推移のみを考慮した場合で、被験者のタップ入力から得られた 959 個のデータと比較したところ、目的の曲が指定順位以内に入った割合は表 9.1 および表 9.2 のとおり。

ここで、各被験者別における順位を表 9.3 および表 9.4 に示す。MIDI による DB で音高推移のみを比較した場合は、 $P < 0.05$ より、タップのリズムも併用して評価した場合に比べて有意差が見られた。

また、楽曲 DB に含まれている 100 曲の MIDI データからタップ入力のデータを作成した場合、タップのリズムのみを評価した場合、音高推移のみを評価した場合、両方を同時に評価した場合の全てにおいて、100%の割合で、目的の曲を 1 位にすることができていた。楽曲波形による DB および MIDI による DB において、指定した順位に入った割合を表 9.5 に示す。

LCS(最長共通部分列問題 - Longest Common Subsequence) により、音高推移のみを比較した時の一致率を表 9.6 に示す。

9.3 考察

実験結果より、MIDI データからタップ入力データを生成した場合は、リズムのズレのみを考慮した場合および音高推移のみを考慮した場合において正確な結果を出していることも判明した。

全体的には、MIDI による楽曲 DB と音高推移とタップのリズムの両方を比較することにより、聴きたい楽曲が 5 位以内に入る確率を上げることが出来た。しか

しながら、特に音高推移のみを比較した場合において、被験者によって選曲の際のスコア順位に差が出ていることがわかった。

音高推移を入力させる場合、被験者が楽曲の音高を聞き分けることができ、それを入力することができた場合は良い結果を出すことができた。しかしながら、そうでない場合は目標の曲のスコア順位が下がってしまうなど、音高推移が結果に悪影響をおよぼしている可能性もあることがわかった。特に、目標の曲が5位以内に入った確率が60%以上になる場合において、リズム・音高の両方を考慮した場合に確率が良くなる被験者とリズムのズレのみを考慮した場合に確率が良くなる被験者、どちらでもない被験者に分かれていることがわかった。これは、被験者が音高を識別するのが難しかったことによりリズムを重視してタップを行った場合に、リズムのズレのみを考慮した比較方法のほうが確率が良くなっている可能性がある。また、リズムを正確にタップするのに慣れていない被験者の場合、聴きたい曲が上位に上がる確率が下がっていると考えられる。

また、音高推移のみを正しく入力しても、リズムのずれによりスコアが正確に計算されない可能性がある。これは、このアルゴリズムおよび提案手法のアルゴリズムは、入力のタップと楽曲DBの発音のペアを、タップの時刻のみを用いて結びつけているためであり、リズムがずれてしまった場合は、音高推移が正しく選曲に反映されない可能性があると考えられる。

よって、楽曲比較アルゴリズムを改良する場合、音高推移と楽曲の発音リズムの両方を適切に利用する、または、ユーザーの熟練度によりアルゴリズムを変えるなどの工夫を行うことにより、より正確な結果を出すことが出来ると考えられる。

聴きたい曲が5位以内となった割合が、MIDIによるDBに比べて楽曲波形によるDBの場合は大幅に低下している。これは、楽曲波形からのDBの作成時に誤った発音リズム・周波数を数多く検出したことによるものと考えられる。そのため、楽曲波形を解析するアルゴリズムを改良する必要があると考えられる。

このシステムにより、音高推移とリズムを同時に入力させる場合、被験者によっては慣れやある程度の訓練が必要となってくる可能性が高い。しかしながら、それらに慣れている場合、従来手法よりも高速で選曲できるようになる可能性がある。

表 9.1: 各曲の入力における, 目的の曲が指定順位以内に入った割合 (MIDI)(%)

被験者	比較内容	1 位	5 位以内	10 位以内	20 位以内	30 位以内
1	リズムのズレ	9.2	60.8	75.0	86.7	91.7
	音高推移	28.3	60.0	68.3	80.0	91.7
	両方	15.0	73.3	81.7	85.8	92.5
2	リズムのズレ	13.3	55.0	65.8	78.3	85.0
	音高推移	18.3	51.7	67.5	80.0	89.2
	両方	19.2	65.0	73.3	83.3	87.5
3	リズムのズレ	0.0	22.7	29.4	45.4	54.6
	音高推移	0.0	14.3	22.7	38.7	55.5
	両方	0.0	18.5	31.1	41.2	56.3
4	リズムのズレ	5.8	36.7	49.2	64.2	73.3
	音高推移	7.5	21.7	30.8	55.0	65.8
	両方	2.5	27.5	46.7	59.2	68.3
5	リズムのズレ	2.5	20.8	35.8	60.0	70.0
	音高推移	4.2	15.8	25.8	39.2	50.0
	両方	3.3	20.0	31.7	50.8	57.5
6	リズムのズレ	8.4	48.2	59.7	73.0	79.6
	音高推移	80.8	90.8	92.5	94.2	96.7
	両方	70.0	89.2	95.0	96.7	97.5
7	リズムのズレ	13.3	65.8	83.3	90.8	91.7
	音高推移	86.7	91.7	94.2	95.8	97.5
	両方	66.6	90.0	94.1	95.8	95.8
8	リズムのズレ	8.4	48.2	59.7	73.0	79.6
	音高推移	48.3	72.5	80.0	87.5	89.2
	両方	43.3	74.2	81.7	89.2	90.8
平均	リズムのズレ	15.8	59.2	65.8	80.0	87.5
	音高推移	34.3	52.3	60.3	71.3	79.5
	両方	27.5	57.2	66.8	75.2	80.7

表 9.2: 各曲の入力における, 目的の曲が指定順位以内に入った割合 (楽曲波形)(%)

被験者	比較内容	1 位	5 位以内	10 位以内	20 位以内	30 位以内
1	リズムのズレ	1.7	9.2	17.5	32.5	47.5
	音高推移	4.2	16.7	26.7	37.5	44.1
	両方	5.0	15.8	26.7	37.5	51.7
2	リズムのズレ	7.5	15.8	30.8	52.5	63.3
	音高推移	5.0	15.0	23.3	43.3	57.5
	両方	2.5	16.7	30.0	50.0	62.5
3	リズムのズレ	0.0	2.5	10.1	23.5	42.9
	音高推移	0.0	5.0	15.1	22.7	35.3
	両方	0.0	4.2	8.4	23.5	36.1
4	リズムのズレ	0.8	9.2	23.3	35.8	51.7
	音高推移	2.5	7.5	11.7	26.7	36.7
	両方	2.5	6.7	14.2	27.5	41.7
5	リズムのズレ	4.2	14.2	28.3	47.5	60.8
	音高推移	5.8	15.8	25.8	42.5	55.0
	両方	5.0	15.0	25.8	40.0	55.8
6	リズムのズレ	2.5	10.0	15.8	39.2	52.5
	音高推移	6.7	26.7	39.2	57.5	67.5
	両方	9.2	22.5	33.3	51.7	63.3
7	リズムのズレ	10.0	14.2	28.3	50.8	64.1
	音高推移	16.7	38.4	56.7	68.3	73.3
	両方	13.3	35.8	51.7	65.8	76.7
8	リズムのズレ	0.0	5.8	17.5	32.5	40.8
	音高推移	10.0	17.5	26.7	38.3	55.0
	両方	6.7	12.5	25.8	36.7	47.5
平均	リズムのズレ	3.3	10.1	21.3	39.2	52.9
	音高推移	6.4	17.8	28.2	42.1	53.0
	両方	5.5	27.0	41.5	54.3	64.2

表 9.3: 被験者別平均順位 (MIDI)

被験者	リズム	音高推移	両方
1	9.23	10.76	8.60
2	13.63	12.25	10.47
3	33.27	33.67	33.42
4	19.84	28.80	23.71
5	25.31	35.11	31.67
6	7.90	3.78	3.81
7	12.05	3.38	4.13
8	13.49	9.60	8.49
平均	16.84	17.16	15.54
σ^2	76.1	174	148
$P(\text{対両方})$	0.46	0.048	

表 9.4: 被験者別平均順位 (楽曲波形)

被験者	リズム	音高推移	両方
1	36.80	36.88	36.23
2	30.85	32.25	28.84
3	46.20	43.53	47.10
4	37.01	41.04	40.69
5	33.84	33.33	35.58
6	34.32	25.33	28.69
7	29.12	19.28	18.88
8	42.02	32.22	38.06
平均	32.98	36.27	34.26
σ^2	62.7	31.8	74.7
$P(\text{対両方})$	0.26	0.24	

表 9.5: MIDI から入力した場合において、目的の曲が指定順位以内に入った割合 (%)

DB	比較内容	1 位	5 位以内	10 位以内	20 位以内	30 位以内
波形	リズムのズレ	8.0	19.0	28.0	49.0	59.0
	音高推移	17.0	29.0	42.0	58.0	67.0
	両方	15.0	29.0	43.0	59.0	70.0
MIDI	リズムのズレ	100.0	100.0	100.0	100.0	100.0
	音高推移	100.0	100.0	100.0	100.0	100.0
	両方	100.0	100.0	100.0	100.0	100.0

表 9.6: LCS による音高推移一致率 (%)

被験者	一致率
1	74.6
2	76.5
3	73.1
4	60.6
5	77.1
6	88.9
7	89.9
8	81.6
平均	77.8
σ	8.76×10^{-2}

第10章 おわりに

本研究では、スマートデバイスで画面のタップのリズム・強さ・位置によって楽曲を選曲するシステムを提案し、実装した。

様々な条件で評価を行ったところ、MIDIのDBから選曲した場合に、入力データが正確であれば100%選曲できるが、被験者が手動でタップ入力を行った場合は、最良で5位以内に入る確率が91.7%となった。また、波形データのDBから選曲を行った場合は、入力データが正確であれば、最良で5位以内に入る確率が17.0%となり、被験者が手動でタップ入力を行った場合は同じく38.4%という結果となった。

被験者が入力したデータの音高推移の正解データとの一致率は77.8%であることから、これらの結果は、これは、楽曲を比較するアルゴリズムや楽曲波形から楽曲DBを作成するシステムの改良、ユーザーの慣れや訓練によって徐々に良くなっていくものと考えられる。

楽曲の波形から楽曲DBを作成する場合、ボーカルを正確に除去し、正確に発音タイミングおよびピッチを取得しなければならない。そのため、正確にそれを行う方法を検討することが今後の課題となる。また、このシステムでは、ユーザーが曲を指のタップ操作で表現することに慣れていることが前提とされており、リズムや音感等の能力により、結果が変わっている可能性がある。そのため、ユーザーのどの能力が選曲の精度に影響するのかを調査する必要がある。また、リズムや音感が良くないユーザーでも同じような感覚でより正確に選曲出来る方法も検討したい。

参考文献

- [1] Cristina de la Bandera, Ana M Barbancho, Lorenzo J Tardón, Simone Sammartino, and Isabel Barbancho. Humming method for content-based music information retrieval. In *ISMIR*, pp. 49–54, 2011.
- [2] 石山英貴, 高橋伸, 田中二郎. コマンドリズムを用いたタップ入力による携帯端末操作手法. 情報処理学会 インタラクシオン 2013, pp. 270–277, feb 2013.
- [3] 池谷直紀, 服部正典, 梅木秀雄, 大須賀昭彦. リズム入力インタフェース「タタタタップ」による大規模音楽検索 (一般セッション, <特集> インタラクシオン・メディアおよび一般). 情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告, Vol. 2005, No. 52, pp. 27–33, may 2005.
- [4] Rui Cai, Chao Zhang, Lei Zhang, and Wei-Ying Ma. Scalable music recommendation by search. In *Proceedings of the 15th International Conference on Multimedia*, MULTIMEDIA '07, pp. 1065–1074, New York, NY, USA, 2007. ACM.
- [5] 野口敦弘, 納富一宏, 斎藤恵一. 自己組織化マップを用いたタッチスクリーンによるリズム認証手法. バイオメディカル・ファジィ・システム学会誌, Vol. 15, No. 1, pp. 31–39, jun 2013.
- [6] Zafar Rafii and Bryan Pardo. Music/voice separation using the similarity matrix. In *ISMIR*, pp. 583–588, 2012.
- [7] Zafar Rafii and Bryan Pardo. Online repet-sim for real-time speech enhancement. In *ICASSP*, pp. 848–852. IEEE, 2013.
- [8] Alain De Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, Vol. 111, p. 1917, 2002.
- [9] marsyas. <http://marsyas.info>.
- [10] FFmpeg. <http://www.ffmpeg.org/>.
- [11] RWC 研究用音楽データベース. <https://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>.

- [12] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一. RWC 研究用音楽データベース : ポピュラー音楽データベースと著作権切れ音楽データベース. 情報処理学会研究報告. [音楽情報科学], Vol. 2001, No. 103, pp. 35–42, oct 2001.

謝辞

本研究にあたり，本研究室の角田博保先生，赤池英夫先生には二年間を通じて様々にご指導いただきました．また，様々にご助言，ご協力をいただき，実験にご協力いただいた研究室の皆様にご心より御礼申し上げます．